Auditory steady-state responses as neural correlates of loudness growth

Maaike Van Eeckhoutte, Jan Wouters, Tom Francart

KU Leuven, Department of Neurosciences, ExpORL

Corresponding author: Maaike Van Eeckhoutte

Herestraat 49, box 721, B-3000 Leuven, Belgium

maaike.vaneeckhoutte@med.kuleuven.be

jan.wouters@med.kuleuven.be

tom.francart@med.kuleuven.be

October 27, 2016

Abstract

The aim of this study was to find an objective estimate of individual, complete loudness growth functions based on auditory steady-state responses. Both normal-hearing and hearing-impaired listeners were involved in two behavioral loudness growth tasks and one EEG recording session. Behavioral loudness growth was measured with Absolute Magnitude Estimation and a Graphic Rating Scale with loudness categories. Stimuli were sinusoidally amplitude-modulated sinusoids with carrier frequencies of either 500 Hz or 2000 Hz, a modulation frequency of 40 Hz, a duration of 1 s, and presented at intensities encompassing the participants' dynamic ranges. Auditory steady-state responses were evoked by the same stimuli using durations of at least 5 minutes. Results showed that there was a good correspondence between the relative growth of the auditory steady-state response amplitudes and the behavioral loudness growth responses for each participant of both groups of listeners. This demonstrates the potential for a more individual, objective, and automatic fitting of hearing aids in future clinical practice.

Keywords

Loudness growth functions; Auditory steady-state responses; Fitting of hearing aids; Objective measure

Highlights

- Auditory steady-state response growth was correlated with behavioral loudness growth
- Correlation was found for both normal-hearing and hearing-impaired participants
- Potential for more objective and automatic fitting of hearing aids

Abbreviations

- ABR: Auditory Brain Stem Response
- AME: Absolute Magnitude Estimation
- ASSR: Auditory Steady-State Response
- DPOAE: Distortion-Product Otoacoustic Emissions
- EEG: Electroencephalogram
- GRS: Graphic Rating Scale
- HI: Hearing-Impaired
- MSE: Mean Square Error
- NH: Normal-Hearing
- OAE: Otoacoustic Emissions

1 Introduction

Loudness growth functions characterize the relation between sound intensity and loudness (Marks and Florentine, 2011). They are highly listener dependent, and thus offer unique information about the hearing of an individual. To date, most prescription rules for non-linear amplification include some aspects of loudness normalization, i.e. they try to make the loudness of the amplified sounds similar to the loudness for normal-hearing listeners listening to the same sound (Dillon, 2012). However, complete loudness growth functions are usually not measured in clinical practice because the procedures for measuring them are time-consuming, complicated, demand active cooperation of the client, are often perceived as difficult by the client, and large variability across people and measurement techniques have been described (Al-Salim et al., 2010; Elberling, 1999).

Examples of loudness growth measures that were used in the past for fitting hearing aids are LGOB or Loudness Growth in half-Octave Bands (Allen et al., 1990), the IHAFF or Independent Hearing Aid Fitting Forum protocol, also known as the Contour Test (Cox, 1995; Valente and Van Vliet, 1997), and ScalAdapt (Kiessling et al., 1996). In these procedures the client needs to estimate the loudness of different sounds, based on loudness categories ranging from not audible or very soft to uncomfortably loud or too loud. The gain of the hearing aid is adjusted to try to achieve normalized loudness. In the ScalAdapt procedure, loudness growth is measured while the client is wearing the hearing aid and its parameters are adaptively adjusted until the client gives a desired loudness rating.

Loudness categories are perceived as simple and easy to understand for inexperienced participants due to their meaningful labels, and previous experience in loudness scaling has no influence on the loudness judgments (Launer, 1995). Because of these factors categorical loudness scaling is more frequently preferred for clinical practice compared to other loudness growth procedures

(Marks and Florentine, 2011; Launer, 1995), even though their reliability and validity have been questioned (Al-Salim et al., 2010; Elberling, 1999). Many loudness scales in the literature have discrete loudness categories, and the number of categories has been the subject of discussion, with too few categories leading to response biases. In the procedure described by Allen et al. (1990), many participants reported that the number of categories, i.e. 6, was insufficient. As a solution, one can add intermediate response categories without labels (Brand and Hohmann, 2001), or use a continuous visual-analogue scale or a graphic rating scale, which is a visual-analogue scale with categories added as guidelines (Marks and Florentine, 2011; Svensson, 2000).

Other procedures have been described to measure loudness growth such as Absolute Magnitude Estimation (AME). The AME task is a classical method for measuring loudness, often proposed as the most direct and effective method (Hellman and Meiselman, 1990; Marks and Florentine, 2011).

Attempts have been made to find an objective, more automatic, and physiological correlate of loudness growth functions using different kinds of measures. While these measures have sources at different stages of the auditory pathway, at present, it is not fully understood at which stage of the auditory pathway the loudness coding is complete for different stimuli.

Otoacoustic emissions (OAEs) have been assessed as one correlate of loudness growth. OAEs might be practical to use because they are fast to acquire. OAEs are generated by the outer hair cells in the cochlea in response to acoustic stimuli, and can be measured in the ear canal. Thus, this approach is based on the assumption that the perception of loudness is mainly determined at the level of the outer hair cells, while it is likely that loudness is also affected by other auditory processes for which OAEs are insensitive, such as processes at the level of the inner hair cells, synaptic and neural functions, and central auditory processes. Loudness growth has been linked to both distortion-product otoacoustic emissions (Neely et al., 2003; Müller and Janssen, 2004; Rasetshwane et al., 2013; Thorson et al., 2012) and tone-burst otoacoustic emissions (Epstein and Florentine, 2005; Epstein and Silva, 2009; Silva and Epstein, 2010, 2012) for normal-hearing and hearing-impaired participants. However, correlations between loudness and DPOAEs were only found if multiple linear regression analyses utilizing the entire DPOAE input-output function were used instead of individual DPOAE input/output function parameters. Furthermore, the DPOAE data showed large inter-subject variability, with good agreement only with group medians. Other disadvantages have been described for the use of OAEs. First, the use of OAEs is limited to individuals with mild-to-moderate hearing loss, since OAEs are absent for individuals with greater degrees of hearing loss. Second, the reliability of OAE measurements is affected by several factors, such as calibration errors, probe-tip placement, recording instruments, and environmental noise (Keppler et al., 2010). Third, at frequencies near the ear-canal resonance, such as 4 kHz, loudness estimates using OAEs are unreliable (Silva and Epstein, 2010, 2012).

Another possible correlate of loudness that has been extensively investigated is the auditory brain stem response (ABR), an auditory evoked potential. Correlations between loudness and ABR amplitude or latency growth functions were either low and not significant for all participants (Wilson and Stelmack, 1982), or were only significant when averaged results were used across participants or test trials. For normal-hearing participants, most of the studies do not show a direct link between the ABR and loudness growth (Babkoff et al., 1984; Darling and Price, 1990; Davidson et al., 1990; Pratt and Sohmer, 1977; Serpanos et al., 1997). Davidson et al. (1990) analyzed the ABR wave V amplitude, and Serpanos et al. (1997) the ABR wave V latency, while the other studies investigated the amplitudes and latencies of multiple waves (I-VI). Serpanos et al. (1997) found a relation between the ABR wave V latency and loudness growth for participants with flat hearing loss. However, there was no such relation for participants with a sloping hearing loss. This study also used averaged group results. Furthermore, the use of ABR has two major disadvantages. First, the waveform of the ABR is often subjectively labeled. Second, there is a lack of frequency specificity, since ABRs are often evoked by click stimuli. To address these problems, Silva and Epstein (2010, 2012) developed an automatic analysis and segmentation method to use with ABRs evoked by 1- and 4- kHz tone-burst stimuli and reported reliable loudness growth estimates if residual noise levels, i.e. the amount of noise left in the final averaged waveform that affects the ABR amplitude estimation, are controlled. Residual ABR noise levels were estimated through the weighted nonstationary fixed-multiple-point (WNS FMP) statistic, and used as weights in a subsequential non-linear fit with a polynomial or with shifted versions of the INEX function. In summary, mixed results were reported for the relation between ABRs and loudness growth, with many studies finding a lack of correspondence.

Since a lack of correspondence with loudness growth was often found with OAEs and ABRs, loudness may not be fully determined by neural activity at the level of the outer hair cells or the brain stem. Evidence regarding a cortical basis of loudness was suggested by Heinz et al. (2005) and described by Thwaites et al. (2016). Heinz et al. (2005) found that the auditory nerve rate functions of cats with noise-induced hearing loss were inconsistent with the hypothesized neural correlates of loudness recruitment. Thwaites et al. (2016) investigated the location of cortical entrainment to two realistic models of sound magnitude, i.e. the instantaneous and short-term loudness models. Instantaneous loudness is assumed to be the loudness after transformation at peripheral levels, and is already represented in the brain but not yet available to conscious perception, while short-term loudness is formed by running temporal integration of the instantaneous loudness. The location of cortical entrainment to instantaneous loudness was found in Heschl's gyrus. It was suggested that it is moved or copied to the dorsal lateral sulcus and from there back to Heschl's gyrus. Cortical entrainment to the short-term loudness was found in both the dorsal lateral sulcus and superior temporal sulcus.

Correlations between loudness growth and objective measures based on sources further in the auditory pathway than the outer hair cell, auditory nerve, or brain stem have also been investigated. Madell and Goldstein (1972) found high correlations across participants between the peak-to-peak middle-latency response amplitudes and loudness estimates, with correlation coefficients of 0.94, 0.85, and 0.75 for P0-Na, Na-Pa, and Pa-Nb, respectively. However, no significant correlations were found for individual participants. Pratt and Sohmer (1977) found no correlations between loudness estimates and the cortical responses P1-3 evoked by a series of click stimuli with peak energy in the 3-5 kHz range. They proposed that the loudness estimate is likely determined by neural activity that is not registered by the recording technique and that another set of neural parameters might be required to estimate loudness.

Several fMRI studies support the hypothesis that the loudness percept is not complete before the level of the auditory cortex. For normal-hearing participants, significant correlations between loudness and the extent and the magnitude of cortical activation or the fMRI blood oxygen level dependent signal (BOLD-signal) were found at the auditory cortices, but not at any lower sources of the auditory pathway such as the inferior colliculus or the medial geniculate bodies (Hall et al., 2001; Röhl et al., 2011; Röhl and Uppenkamp, 2012; Uppenkamp and Röhl, 2014). For participants with a high-frequency hearing loss, steeper growth in the magnitude of the cortical responses with sound intensity was found for high-frequency FM-tones (4-8 kHz) than for low-frequency FM-tones (0.5-1kHz), which was interpreted as a correlate of the psychoacoustic effect of loudness recruitment (Langers et al., 2007).

The auditory steady-state response (ASSR) might be a good objective correlate of loudness growth. The ASSR is a stationary neural response to a periodic stimulus, and can be detected in the electroencephalogram (EEG) (Picton, 2011). The ASSR is frequency-specific and can be measured fully objectively through statistical tests. It also has the potential to be measured automatically and quickly. Instead of ear-by-ear or frequency-by-frequency testing it is possible to evoke ASSRs using multiple simultaneous stimuli (Ishida and Stapells, 2012; Lins and Picton, 1995).

There are several reasons why the ASSR might be a useful tool for estimating loudness growth functions. First, the amplitude of the ASSR grows nonlinearly with intensity (e.g., Lins and Picton, 1995; Picton et al., 2007), as do loudness growth functions for normal-hearing and hearing-impaired participants (e.g., Moore, 2007). Second, for hearing-impaired participants the ASSR amplitude growth is steeper than for normal-hearing participants (Dimitrijevic et al., 2002; Picton et al., 2005). Although no comparisons with loudness were made, the steeper growth in ASSR amplitude was called "physiological recruitment", since it resembles the loudness recruitment phenomenon of hearing-impaired participants who experience an abnormally rapid growth in loudness with increasing intensity (e.g., Moore, 2012).

The main neural sources of the ASSR are determined by the modulation frequency of the stimulus. A modulation frequency around 80 Hz is frequently used, since for this modulation frequency the ASSR mainly originates from the brain stem and is therefore less affected by sleep and sedation, which makes it suitable to use with young children. Using modulation frequencies around 80 Hz, Ménard et al. (2008), Zenker Castro et al. (2008), and Emara and Kolkaila (2010) did find correlations between loudness growth and ASSR amplitude growth, but Israelsson et al. (2015) did not recommend the use of this ASSR amplitude growth function for fitting nonlinear hearing aids due to the high variability of the amplitude growth functions among participants. However, in this study no comparisons to behavioral loudness growth functions were made.

The ASSR evoked by stimuli with a modulation frequency around 40 Hz rather than 80 Hz might be a better correlate of loudness growth, since the largest response amplitudes and signal-to-noise ratios are found with a modulation frequency of 40 Hz for adult awake participants. This ASSR has a clear dominant source at the primary auditory cortex, although contributions of subcortical sources have been described, such as the thalamus and midbrain (e.g., Reyes et al., 2005; Steinmann and Gutschalk, 2011).

The aim of this study was to investigate the relation between ASSR amplitude growth functions, evoked by stimuli with a modulation frequency of 40 Hz, and loudness growth functions. To assess whether the ASSR might be useful for individual fitting of hearing-impaired adults, the behavioral and ASSR results were compared for each individual.

2 Material and methods

2.1 Participants

Two groups of participants were tested. All participants provided their informed consent in accordance with the declaration of Helsinki, and the project was approved by the ethical committee of the University Hospital of Leuven (UZ Leuven). None of the participants had prior experience with loudness growth tasks. All participants were native Dutch speakers. The participants' travel expenses were reimbursed.

The first group consisted of 15 normal-hearing participants (8 women, 7 men) with a mean age of 22 $\pm$ 3 years. Their normal hearing was confirmed with pure tone audiometry for the test ear with a Madsen Electronics Orbiter 922 audiometer and TDH-39 headset. All participants had thresholds of 20 dB HL or better for all octave frequencies between 0.125 and 8 kHz, with the exception of one participant who had a threshold of 25 dB HL at 8 kHz.

The second group consisted of 15 hearing-impaired participants (6 women, 9 men), with a mean age of 65 $\pm$ 15 years. As assessed by pure tone audiometry (air and bone conduction), 13 participants had a sensorineural hearing loss and 2 participants had a mixed hearing loss. Both groups received an otoscopic examination before each test session to ensure non-obstructed ear canals. The details of the hearing-impaired participants are given in Table 1.

Since there is a consensus in literature that age (for adults) does not affect the 40-Hz ASSR for amplitude-modulated stimuli, we can assume that the different ages of the two groups will not confound our results (e.g., Goossens et al., 2016; Grose et al., 2009).

The handedness of the participants was assessed with the Edinburgh Handedness Inventory (Oldfield, 1971). In the group of normal-hearing participants, 10 were right-handed, 2 were left-handed, and 3 were ambidextrous, and in the group of hearing-impaired participants, 13 were right-handed, 1 was left-handed, and 1 was ambidextrous. We did not exclude any ambidextrous or left-handed participants, because their results were similar to those for the other participants. The participants were asked whether they had tinnitus, and only participants who did not have tinnitus or only a soft negligible tinnitus were considered for participation.

2.2 Stimuli and apparatus

Testing was performed in a soundproof booth. The ASSRs were recorded in an additionally electromagnetically shielded booth. Sinusoidally amplitude-modulated (SAM) sinusoids were presented monaurally through an Etymotic Research ER-3A insert ear phone, connected to an RME Hammerfall DSP Multiface II sound card. The stimuli were created in Matlab R2013a (The MathWorks, Inc., Natick, MA) and are described by the following formula:

$$y(t) = (0.5 + 0.5 * \sin(2\pi f_m t)) * \sin(2\pi f_c t) \quad (1)$$

with $y(t)$ the stimulus amplitude over time, f_c the carrier frequency of 500 Hz or 2000 Hz, and f_m the modulation frequency of 40 Hz. The stimuli were calibrated using a 2cc Brüel & Kjær coupler type 4152. The stimulus intensity is described below.

The stimulus duration was 1 s in the behavioral loudness tasks, and 5 to 10 minutes for ASSR recordings. A behavioral stimulus duration of 1 s was chosen in order to prevent temporal integration effects on the loudness judgments (Marks and Florentine, 2011). Loudness adaptation effects with a stimulus of several minutes occur only for the SAM 2000 Hz stimulus at low levels (Van Eeckhoutte et al., 2015). Preliminary results concerning ASSR amplitude changes over time indicated no meaningful adaptation effects.

The software platform APEX3 (Francart et al., 2008) was used for the behavioral experiments. For the ASSR recordings, the software platform for the Recording and analysis of Brain responses to Auditory stimulation (RBA, Hofmann and Wouters, 2012), was used. The signal sampling rate was 32 kHz. The EEG was recorded with the ActiveTwo System Software (Biosemi) using a recording sampling rate of 8192 Hz. A head cap consisting of 64+2 Ag/AgCl active scalp electrodes was mounted on the head in accordance with the standard 10-20 electrode position system (see Figure 1).

The left ear of the participants was chosen for stimulation with the exception of 4 participants. For these participants, the right ear was tested because of an obstructed ear canal in the left ear or in case of a unilateral hearing loss with a normal-hearing left ear. The contralateral ear was plugged to minimize background noise and other distractions.

2.3 Behavioral loudness tasks

Similar protocols were used for both groups of participants. First, an estimate of the dynamic range was obtained for hearing-impaired participants only.

Then, behavioral loudness growth was measured with two tasks: Absolute Magnitude Estimation (AME), and a Graphic Rating Scale (GRS).

Dynamic range estimation First, the detection threshold for each stimulus was measured with an adaptive, one-interval, three-alternative forced-choice (3AFC) procedure without feedback. The participants had to choose one out of three intervals on a computer screen that were lighted up consecutively as the interval containing the stimulus. The level of the stimulus was adjusted based on a two-down, one-up rule, converging on 71% correct. The step sizes were 10, 5, and 2 dB after 0, 1, and 3 reversals, respectively. The task ended after 6 reversals, and the threshold was calculated as the mean level at the last 6 trials.

Second, the maximum acceptable level of each stimulus was measured with an adjustment procedure. The participant was asked to indicate the loudness of the stimuli on the GRS. The participant could choose any position on the scale, with the loudness categories serving only as guidelines. The intensity of the first stimulus was presented slightly above threshold. The stimulus intensity was increased by the experimenter until the participant indicated that the loudness of the stimulus corresponded to "very loud, but still tolerable". The maximum possible level was 115 dB SPL. The experimenter could increase the stimulus intensity with a step size of 1, 2, 5, 10, or 20 dB, and this was depending on the feedback of the participant. The larger step sizes were only used at lower levels.

Behavioral measures of loudness growth Two loudness growth tasks were administered, which both followed the same underlying procedures. The stimuli were always presented between the threshold and the maximum acceptable level. For normal-hearing participants, all levels between 16 and 88 dB SPL with a step size of 6 dB were used, while for hearing-impaired participants, the step sizes were chosen depending on the dynamic ranges in order to have enough data points (using a target number of 15-20 data points). A pseudorandom order of presentation was used, with the constraint that the maximum level difference between two successive stimuli never exceeded half of the participant's dynamic range. This reduces context effects caused by the tendency of participants to judge the loudness of a stimulus relative to the previous stimulus (Brand and Hohmann, 2001). The starting level was 40 dB SPL for the normal-hearing participants or the midpoint of the dynamic range for the hearing-impaired participants. For each carrier frequency tested, there was a training and test phases. For the normal-hearing participants, both training and test phase consisted of 3 repetitions of each level, while for the hearing-impaired participants, the training phase consisted of only 1 repetition of each level to save measurement time, but the test phase again consisted of 3 repetitions.

For the first loudness growth task, *AME*, the participants were instructed to rate the loudness of each stimulus by typing a number. They were free to choose any positive number, even decimals and fractions, with zero meaning that the stimulus was inaudible. The participants were explicitly instructed that there is always an infinite range of numbers between two numbers, that it was allowed to use the same number several times, and that the answers could never be wrong. No examples were given in advance. The AME task was always conducted first.

For the second loudness growth task, *GRS*, the scale shown in Figure 2 was used. The procedure was exactly the same as for the AME task, but the participants had to choose a position on the scale instead of judging loudness with numbers. Any position on the scale could be chosen, including between the loudness categories. The loudness categories only served as a guideline. The participants clicked with a computer mouse on a position on the scale, which was shown on the computer screen. The software coded the chosen position as a number between 0 (corresponding to "Inaudible") and 1 (corresponding to "Unbearable"). The participants were explicitly instructed that one region on the scale could be chosen more often than another region, and that an answer could never be wrong.

2.4 EEG recordings for ASSR growth functions

For the normal-hearing participants, the ASSR measures were generally obtained on the same day or within two weeks of the behavioral tasks. For all hearing-impaired participants, the ASSR recordings took place directly after the behavioral loudness growth tasks.

Before the start of the EEG recordings, a stimulus with the highest level was briefly presented to the participant. If the participant felt that it would be too loud to listen to for 5 to 10 minutes, a lower level was chosen, which was one of the levels used in the behavioral loudness growth tasks. Up to 8 levels were chosen and those were also used in the behavioral loudness growth tasks.

During the EEG recordings, the participants sat in a comfortable chair or lay down on a bed, and were instructed to relax as much as possible. A subtitled, silent video that could be chosen in advance was presented to prevent participants from falling asleep. The stimuli were presented consecutively with increasing stimulus level while EEG recordings were made. For stimulation at low levels the EEG was often recorded for 10 minutes, or the recordings were terminated if the real-time monitor indicated that a significant response was reached (but with a minimum of 5 minutes). The real-time monitor combined the EEG-signals of relevant electrodes by averaging. The number of epochs used in the analysis was incremented step-by-step, and at each step a Hotelling t^2-test determined the significance of the response. As a correction for repeated testing, at each test step the critical value was adjusted to ensure a fixed false alarm rate of 5%. The real-time monitor was not used for analysis of the data. The two carrier frequencies were presented alternately, to prevent possible adaptation effects (Van Eeckhoutte et al., 2015). Breaks were given depending on needs, with at least two breaks per participant.

The data were analyzed offline using Matlab R2013a (The MathWorks, Inc., Natick, MA). The raw data were filtered using a second-order butterworth high-pass filter with a cut-off frequency of 2 Hz. The EEG data were then converted into epochs of 1.024 s. The 5% of epochs with the highest peak-to-peak amplitudes were considered as artifacts and rejected. The outcome measure was the response amplitude determined with the Hotelling t^2-test after Fast Fourier Transform (FFT), which uses both response amplitude and phase obtained from the complex frequency bin at the modulation frequency. The significance level was set at $\alpha = 0.05$. In all further analyses, only significant ASSR amplitudes are considered. A significant ASSR amplitude means that the complex response bins at the modulation frequency were significantly different than the spontaneous measured EEG activity. Only recordings from active electrodes for which 80% of the participants showed significant ASSR amplitudes were used. All electrodes were referenced to Cz. Of these electrodes, only bilateral pairs of electrodes were considered, as well as midline electrodes. This resulted in the following electrode selection: P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, PO7, PO8, PO3, PO4, O1, O2, Iz, Oz, POz, and Pz (see Figure 1). The final ASSR amplitude for each carrier frequency and participant was the average of the significant ASSR amplitudes of the selected electrodes. Nevertheless, very similar results could be obtained when using other electrode selections, such as only one electrode (e.g., P10), or after a Denoising Source Separation (DSS) analysis (de Cheveigné and Simon, 2008).

2.5 Comparison of measures

The three measures (AME, GRS, and ASSR) for each participant and each carrier frequency were transformed to allow them to be compared. Mean square errors were used to statistically investigate the differences between the three measures. The statistical analyses were performed using R (R Core Team, 2014).

Data transformation For behavioral responses with multiple responses at a certain stimulus level, the mean was taken. Then, the equation described in

Silva and Epstein (2010, 2012) was used for the transformation of the responses. The logarithm of the response was subtracted from the logarithm of each response to obtain zero-mean curves. Specifically, for each participant, measure (AME, GRS, and ASSR), and carrier frequency we calculated:

$$D_i = \log_{10} R_i - \frac{1}{N}\sum_{j=1}^{N} \log_{10} R_j \tag{2}$$

with D_i the transformed response for a given level i, R_i and R_j the response for a given level i or j, and N the total number of levels.

Subsequently, to make the data easier to interpret, all measures were transformed back to GRS values. The final transformed values were obtained by elevating 10 to the power of the addition of Di and the mean of the transformed GRS responses for the particular carrier frequency and participant that was tested:

$$F_i = 10^{(D_i + \mu_{GRS})} \tag{3}$$

with F_i the final transformed value, D_i the transformed value after Equation (2), μ_{GRS} equal to $\frac{1}{N}\sum\limits_{j=1}^{N} \log_{10} R_j$ and R_j the GRS response of a particular participant and carrier frequency, wherein j varies from 1 to the total number of levels N.

Statistical comparison of measures After transformation, each of the measures (AME, GRS, and ASSR) was compared to each other measure by calculating the mean square error (MSE) between the two curves. MSEs were used because they are robust and do not assume a linear function.

For statistical analysis, outliers were removed for each MSE comparison and group of participants based on the median absolute deviation or MAD-median rule (Wilcox et al., 2013). An MSE value X_p of a participant p was considered an outlier if

$$\frac{|X_p - M|}{MAD/0.6745} > 2.24 \tag{4}$$

with M the median of the MSE values across a group of participants for a given MSE comparison, and MAD the median of $|X_1 - M|, ..., |X_n - M|$, with n the number of participants. Over all MSE comparisons, 13 outliers out of 90 values (15 participants x 2 carrier frequencies x 3 MSE comparisons) were removed for the normal-hearing participants, and 10 out of 90 were removed for the hearing-impaired participants.

A linear mixed-effects model was used that included the following factors: the MSE comparison (AME-GRS, AME-ASSR, or GRS-ASSR), Carrier Frequency (500 or 2000 Hz), and Participant Type (normal-hearing or hearing-impaired), with MSE Comparison and Carrier Frequency included as repeated-measure factors. Two contrasts were also set. The contrast "Beh-ASSR" was the "Behavioral vs ASSR" contrast, which compared the behavioral MSE comparison (AME-GRS) with the MSE comparisons that also contained ASSR responses (AME-ASSR and GRS-ASSR). The contrast "Diff-ASSR" was the "Differences in ASSR" contrast, which compared the two MSE comparisons that contained ASSR responses, i.e. the AME-ASSR and the GRS-ASSR conditions. Interactions of the factors were also considered and the significance level was set at $\alpha = 0.05$.

3 Results

Example results Examples of responses of a normal-hearing and a hearing-impaired participant are shown in Figure 3. The hearing-impaired participant had a low-frequency hearing loss. Based on visual inspection, the shapes of the growth functions for the AME responses, the GRS responses and the ASSR amplitudes were similar within each participant, for both frequencies. The EEG background noise was always stable across measurements. For the behavioral loudness growth measures, each error bar in Figure 3 indicates the mean $\pm$ standard deviation of all responses given at one stimulus level. Data for the training and test phases for the behavioral responses were combined, since high correlation coefficients were found between the responses for the training and test phases for each combination of behavioral loudness growth measure and carrier frequency for both groups of participants (see also Table 2).

Transformed results Since normal-hearing participants were always presented with the same stimulus levels, the transformed responses for each stimulus level were averaged for each carrier frequency and compared to predictors of two widely used loudness models for normal hearing, which were transformed in the same way (Figure 4). The conversion from sones to categorical units between 0 and 50 was based on Heeren et al. (2013). These values were divided by 50 in order to have comparable data to our GRS data. Overlapping error bars in Figure 4 indicate a good correspondence between measures, and the standard deviations indicate variability across participants.

The first loudness model was the Inflected Exponential or INEX model as described in Marozeau (2011). The model is a modification of the classical power function for loudness growth and can be written as a fifth-order polynomial. Although a sinusoid of 1000 Hz presented binaurally was used in the experiments leading to the INEX model, while 500 and 2000 Hz stimuli with a modulation frequency of 40 Hz were presented monaurally in the current study, there was a reasonably good correspondence with the averaged results of this study. The second loudness model was the model of Moore and Glasberg (1997), which can be implemented for each stimulus separately. A good correspondence was found between the predictions of this model and the averaged responses of the participants for both carrier frequencies.

Individual transformed results are shown in Figures 5 and 6. The figures show the individual results of all hearing-impaired participants for the two carrier frequencies. No responses were obtained from two hearing-impaired participants, HI1 and HI2, for the AME measure. As can be seen in the figures, in many cases steeper growth functions were associated with higher thresholds. Overall, the ASSR amplitudes were close to both behavioral measures on an individual basis. Normal-hearing participants demonstrated results similar to those for hearing-impaired participants when they had low thresholds for one of the carrier frequencies, e.g., participants HI6 and HI10 with a high-frequency hearing loss had nearly normal thresholds at 500 Hz.

Statistical comparison of measures Similar median MSE values over all MSE comparisons were found for all combinations of measures and carrier frequencies. Taking into account only MSE comparisons with ASSRs, MSE values were between 0.010 and 0.016 for normal-hearing participants, and between 0.005 and 0.009 for hearing-impaired participants. Figure 7 shows the MSEs for each group of participants and each MSE comparison, without outlier removal.

The results of the linear mixed-effects model are shown in Table 3. There were significant main effects of Carrier Frequency and Participant Type, and a significant interaction between Carrier Frequency and Participant Type. Post-hoc tests revealed that the 500 Hz carrier frequency was significantly lower for hearing-impaired participants than for normal-hearing participants ($p < 0.001$), and that the 500 Hz carrier frequency was significantly lower than the 2000 Hz carrier frequency for hearing-impaired participants ($p = 0.001$). No other significant effects were found. The p-values were corrected based on Holm's method.

Similar MSE values were obtained when using a more clinically used electrode configuration, e.g., with mastoid electrodes P9 and P10. These MSE values are shown in Table 4.

4 Discussion

Behavioral loudness growth functions measured with two tasks (AME and GRS) and ASSR amplitude growth functions were compared for normal-hearing and hearing-impaired participants. Levels were included that encompassed the participants' dynamic ranges. After transformation of the responses to directly compare the three measures (AME, GRS, and ASSR), good correspondence was found, with median MSE values between 0.005 and 0.016 for MSE comparisons including ASSR values.

To interpret the magnitude of these MSE values, the values can be related to a typical root mean square error on a GRS loudness scale between 0-1 and a loudness scale with loudness categories between 0 and 50 (corresponding to inaudible and too loud) (Brand and Hohmann, 2001). The obtained values correspond to 0.07 and 0.13 on a GRS scale, and to 3.5 and 6.3 on a scale with categorical units, respectively.

The ASSR amplitude growth functions had almost identical shapes as the loudness growth functions measured behaviorally. For example, if loudness recruitment was present in the behavioral loudness growth measures, it was also present in the ASSR amplitude growth functions (see Figures 5 and 6).

No significant differences were found between MSE comparisons, except the significant interaction between carrier frequency and participant type, with slightly better results for the 500 Hz carrier frequency for hearing-impaired participants. Additionally, correlations were calculated between the MSE values and the thresholds of the participants for each carrier frequency. All r-values were low and non-significant, suggesting that the different dynamic ranges of the hearing-impaired participants did not have an influence on the obtained MSE values. Consequently, the loudness estimates obtained from the ASSR amplitudes were good predictors of the loudness estimates obtained using the two behavioral measures, especially for hearing-impaired participants.

The measurement error in the behavioral measure defines the precision of the loudness growth function for each participant. Small MSE values between behavioral loudness growth functions indicated reliable estimates.

Comparison to other ASSR studies Even though large variability among participants has been reported in the literature for both behavioral loudness growth functions and ASSR amplitude growth functions (e.g., Elberling, 1999;

Israelsson et al., 2015), previous studies focused on group averaged results for the comparison of ASSR amplitude growth and behavioral loudness growth (Emara and Kolkaila, 2010; Ménard et al., 2008; Zenker Castro et al., 2008). Good within-subject reliability has been reported for both behavioral loudness growth measures and ASSR amplitudes (Al-Salim et al., 2010; D'Haenens et al., 2008; Robinson and Gatehouse, 1996).

To compare our results to earlier results, we calculated linear correlation coefficients for group data in the same way as Emara and Kolkaila (2010); Zenker Castro et al. (2008), and Ménard et al. (2008), even though we think MSEs are better suited for analyzing the non-linear data in this study. These studies used modulation frequencies in the range of 80-100 Hz. Since these studies used the Contour Test or a category scale to measure loudness behaviorally, we only included the responses for the GRS in the analysis. Emara and Kolkaila (2010) reported linear correlation coefficients between loudness judgments and ASSR amplitudes for normal-hearing participants. Correlation coefficients were $r = 0.55$, $r = 0.62$, and $r = 0.55$ for carrier frequencies of 1000, 2000, and 4000 Hz, without any transformation of the responses. The correlations obtained by doing the same analysis for the data from the normal-hearing participants in this study were $r = 0.73$ and $r = 0.68$, for the 500 and 2000 Hz carrier frequencies, respectively. Zenker Castro et al. (2008) reported correlation coefficients obtained from a multiple regression formula to predict loudness from the ASSR amplitude and the intensity in normal-hearing participants. Correlations between 0.82 and 0.85 were found for carrier frequencies between 500 and 4000 Hz. Correlations obtained from a multiple regression analysis in this study were 0.91 and 0.88 for the 500 and 2000 Hz carrier frequencies, respectively. Ménard et al. (2008) transformed the data by dividing the responses by the maximum response, and found a correlation of $r = 0.90$ (originally reported as an R^2 value of 0.81) between the ASSR amplitudes and loudness for normal-hearing participants. For the data of this study, the correlations were $r = 0.94$ for both the 500 and 2000 Hz carrier frequencies. In summary, the correlation coefficients in this study were consistently slightly higher than the ones reported in the above-mentioned studies, possibly due to the difference in modulation frequency and the corresponding source difference of the responses.

Comparison to OAE and ABR studies using the same analysis Due to the large inter-subject variability, the statistical analysis in this study focused on the MSE values calculated for each participant. Since Silva and Epstein (2010, 2012) used the same transformation (only the first step of our transformation, see equation (2)) and MSE calculation on individual data sets to estimate loudness growth functions with ABRs or OAEs, the MSE values of this study can be directly compared to theirs.

The two behavioral loudness growth measures gave similar results for both normal-hearing and hearing impaired participants. In Silva and Epstein (2010, 2012), median behavioral MSEs were calculated based on Cross-Modality Matching (CMM) and AME, and were 0.08 and 0.12 at 1 and 4 kHz for normal-hearing participants, and 0.03 and 0.01 for hearing-impaired participants, respectively. Calculated with only equation (2), median MSE values (without outlier removal) in this study were for both the normal-hearing and hearing-impaired participants 0.01 and 0.02 for the 500 and 2000 Hz carrier frequency, respectively. Thus, the MSEs in this study were comparable to or slightly smaller than for their hearing-impaired participants.

Silva and Epstein (2010, 2012) found median MSEs between AME scores and ABRs of 0.09 and 0.08 for 1 and 4 kHz tone bursts for normal-hearing participants, and 0.05 and 0.04 for hearing-impaired participants, respectively, obtained with their best method to control residual noise. The median MSEs between tone-burst OAE growth functions and AME growth functions varied between 0.08 and 0.12 for a 1 kHz tone for normal-hearing and hearing-impaired participants. However, the median MSEs varied between 0.41 and 0.95 for a 4 kHz tone. Since median MSE values were between 0.01 and 0.02 in this study, a more accurate objective estimation of loudness growth functions can be obtained with ASSRs than with ABRs or OAEs.

Neural sources Previous studies focused on objective measures with sources from the outer hair cells (OAEs) or the brain stem (ABRs and ASSRs evoked using a modulation frequency of 80 Hz). In this study good correspondence between ASSR growth functions and loudness growth functions was found using a modulation frequency of 40 Hz, which is known to lead to a dominant source in the primary auditory cortex (Picton, 2011). This may indicate that loudness is mediated at a cortical level.

In pilot tests, we also measured ASSR amplitude growth functions for normal-hearing participants using a modulation frequency of 4 Hz to measure even higher cortical sources within the auditory pathway (Picton, 2011). Although a very similar loudness growth curve was obtained using this modulation frequency, the 40 Hz ASSR seems to be preferable, since the EEG background noise becomes larger around 4 Hz leading to worse signal-to-noise ratios.

Applications ASSRs evoked using a modulation frequency of 40 Hz are potentially useful for a more automatic, individual fitting of hearing aids in clinical practice, in addition to threshold and maximum level estimation. In many cases, a combination of behavioral and objective measures will be desired.

If there is no information about the dynamic range, such as for infants, first objective estimates of the threshold and maximum level are needed to prevent stimulating at too low and too high levels. The ASSR is already used for objective estimation of thresholds in clinical practice (Picton, 2011), although thresholds are somewhat higher than behavioral thresholds and correction factors are needed. For estimation of the maximum level, it is possible either to use a fixed maximum stimulus level, or to use another objective measure such as the stapedius reflex threshold. We did not find a consistent saturation of the ASSR amplitudes at the highest stimulus levels.

We demonstrated the feasibility of using 40-Hz ASSRs as an objective measure of loudness growth. The 40-Hz ASSR offers several advantages over other objective measures for estimating loudness growth functions, since it is a frequency-specific method, can be analyzed fully objectively, and has the largest signal-to-noise ratio leading to the shortest recording times in adult awake participants. However, the current protocol is not suitable for clinical practice given the long measurement time. We used a fixed recording time of at least 5 minutes per stimulus level to make sure we had a reliable ASSR amplitude estimate, but our real-time monitor usually indicated a significant response long before the ending of a recording. Often significance was reached after 30 epochs, which correponds to about 31 seconds, for each level and carrier frequency. In clinical practice the test could be stopped once the significance was reached. Furthermore, it might be desirable in clinical practice to use ASSRs evoked by multiple simultaneous stimuli, e.g. testing multiple carrier frequencies in one recording. To avoid possible interaction effects causing a reduction of the ASSR amplitudes (Ishida and Stapells, 2012; Papakonstantinou et al., 2013), the feasibility of the 40-Hz ASSR as an objective measure of loudness growth was investigated with single ASSR recordings in this study. In clinical practice usually only a few electrodes are used to save preparation time, which requires less special and expensive equipment than that used in this study. Many clinics have a 3-electrode set-up available. Very similar results were obtained for other electrode selections, such as only the mastoid electrodes P9 and P10.

5 Conclusion

ASSR amplitudes are feasible to use as an electrophysiological, neural correlate of loudness growth for both normal-hearing and hearing-impaired participants. Behavioral loudness growth functions were measured with Absolute Magnitude Estimation (AME) and a Graphic Rating Scale (GRS). After transformation, the data showed small mean square errors between behavioral loudness growth functions and ASSR amplitude growth functions for two carrier frequencies and both groups of participants. Mean square errors were smaller than for similar studies with otoacoustic emissions and auditory brain stem responses. The 40-Hz ASSR might therefore be a useful tool for more automatic and objective fitting of hearing aids in clinical practice.

Acknowledgments

The authors would like to thank all participants for volunteering. We would also like to thank Marit Schroyen, Julie De Winter, and Ellen Van Avondt for their assistance in recruiting hearing-impaired participants, and Astrid Buelens, Roosmarij Clercx, and Thuur Schilders for their assistance with data collection. We appreciate the help of Dimitar Spirrov in implementing the loudness models. Finally, we are grateful to Brian C.J. Moore and two anonymous reviewers for their valuable remarks to improve the manuscript. The first author was supported by a PhD grant for Strategic Basic Research by the Agency for Innovation by Science and Technology in Flanders (IWT, 131106).

List of Tables

List of Figures

References

Al-Salim, S.C., Kopun, J.G., Neely, S.T., Jesteadt, W., Stiegemann, B., Gorga, M.P., 2010. Reliability of categorical loudness scaling and its relation to threshold. Ear and Hearing 31, 567–578.

Allen, J.B., Hall, J.L., Jeng, P.S., 1990. Loudness growth in 1/2-octave bands (LGOB): a procedure for the assessment of loudness. The Journal of the Acoustical Society of America 88, 745–753.

Babkoff, H., Pratt, H., Kempinski, D., 1984. Auditory brainstem evoked potential latency-intensity functions: A corrective algorithm. Hearing Research 16, 243–249.

Brand, T., Hohmann, V., 2001. Effect of hearing loss, centre frequency, and bandwidth on the shape of loudness functions in categorical loudness scaling. Audiology 40, 92–103.

de Cheveigné, A., Simon, J.Z., 2008. Denoising based on spatial filtering. Journal of Neuroscience Methods 171, 331–339.

Cox, R.M., 1995. Using loudness data for hearing aid selection: The IHAFF approach. The Hearing Journal 48, 10,39–44.

Darling, R.M., Price, L.L., 1990. Loudness and auditory brain stem evoked response. Ear and Hearing 11, 289–295.

Davidson, S.A., Wall, L.G., Goodman, C.M., 1990. Preliminary studies on the use of an ABR amplitude projection procedure for hearing aid selection. Ear and Hearing 11, 332–339.

D'Haenens, W., Vinck, B.M., De Vel, E., Maes, L., Bockstael, A., Keppler, H., Philips, B., Swinnen, F., Dhooge, I., 2008. Auditory steady-state responses in normal hearing adults: a test-retest reliability study. International Journal of Audiology 47, 489–498.

Dillon, H., 2012. Prescribing hearing aid amplification, in: Hearing aids. 2nd edition. Boomerang Press, Turramurra, pp. 286–334.

Dimitrijevic, A., John, M.S., Van Roon, P., Purcell, D.W., Adamonis, J., Ostroff, J., Nedzelski, J.M., Picton, T.W., 2002. Estimating the audiogram using multiple auditory steady-state responses. Journal of the American Academy of Audiology 13, 205–224.

Elberling, C., 1999. Loudness scaling revisited. Journal of the American Academy of Audiology 10, 248–260.

Emara, A.A.Y., Kolkaila, E.A., 2010. Prediction of loudness growth in subjects with sensorineural hearing loss using auditory steady state response. International Advanced Otology 6, 371–379.

Epstein, M., Florentine, M., 2005. Inferring basilar-membrane motion from tone-burst otoacoustic emissions and psychoacoustic measurements. The Journal of the Acoustical Society of America 117, 263–274.

Epstein, M., Silva, I., 2009. Analysis of parameters for the estimation of loudness from tone-burst otoacoustic emissions. The Journal of the Acoustical Society of America 125, 3855–3864.

Francart, T., van Wieringen, A., Wouters, J., 2008. APEX 3: a multi-purpose test platform for auditory psychophysical experiments. Journal of Neuroscience Methods 172, 283–293.

Goossens, T., Vercammen, C., Wouters, J., van Wieringen, A., 2016. Aging affects neural synchronization to speech-related acoustic modulations. Frontiers in Aging Neuroscience 8, 1–16.

Grose, J.H., Mamo, S.K., Hall, J.W., 2009. Age effects in temporal envelope processing: speech unmasking and auditory steady state responses. Ear and Hearing 30, 568–575.

Hall, D.A., Haggard, M.P., Summerfield, A.Q., Akeroyd, M.A., Palmer, A.R., Bowtell, R.W., 2001. Functional magnetic resonance imaging measurements of sound-level encoding in the absence of background scanner noise. The Journal of the Acoustical Society of America 109, 1559–1570.

Heeren, W., Hohmann, V., Appell, J.E., Verhey, J.L., 2013. Relation between loudness in categorical units and loudness in phons and sones. The Journal of the Acoustical Society of America 133, EL314–319.

Heinz, M.G., Issa, J.B., Young, E.D., 2005. Auditory-nerve rate responses are inconsistent with common hypotheses for the neural correlates of loudness recruitment. Journal of the Association for Research in Otolaryngology 6, 91–105.

Hellman, R.P., Meiselman, C.H., 1990. Loudness relations for individuals and groups in normal and impaired hearing. The Journal of the Acoustical Society of America 88, 2596–2606.

Hofmann, M., Wouters, J., 2012. Improved electrically evoked auditory steady-state response thresholds in humans. Journal of the Association for Research in Otolaryngology 13, 573–589.

Ishida, I.M., Stapells, D.R., 2012. Multiple-ASSR interactions in adults with sensorineural hearing loss. International Journal of Otolaryngology 802715, 1–9.

Israelsson, K..E., Bogo, R., Berninger, E., 2015. Reliability in hearing threshold prediction in normal-hearing and hearing-impaired participants using mixed multiple ASSR. Journal of the American Academy of Audiology 26, 299–310.

Keppler, H., Dhooge, I., Maes, L., D'haenens, W., Bockstael, A., Philips, B., Swinnen, F., Vinck, B., 2010. Transient-evoked and distortion product otoacoustic emissions: A short-term test-retest reliability study. International Journal of Audiology 49, 99–109.

Kiessling, J., Schubert, M., Archut, A., 1996. Adaptive fitting of hearing instruments by category loudness scaling (ScalAdapt). Scandinavian Audiology 25, 153–160.

Langers, D.R.M., van Dijk, P., Schoenmaker, E.S., Backes, W.H., 2007. fMRI activation in relation to sound intensity and loudness. NeuroImage 35, 709–718.

Launer, S., 1995. Comparison of different psychoacoustic techniques to measure loudness functions, in: Loudness Perception in Listeners with Sensorineural Hearing Impairment. PhD thesis. Carl von Ossietzky Universität Oldenburg, Germany, pp. 11–24.

Lins, O.G., Picton, T.W., 1995. Auditory steady-state responses to multiple simultaneous stimuli. Electroencephalography and Clinical Neurophysiology 96, 420–432.

Madell, J., Goldstein, R., 1972. Relation between loudness and the amplitude of the early components of the averaged electroencephalic response. Journal of speech and hearing research 15, 134–141.

Marks, L.E., Florentine, M., 2011. Measurement of loudness, part I: methods, problems, and pitfalls, in: Florentine, M., Popper, A.N., Fay, R.R. (Eds.), Loudness. Springer, New York, pp. 17–56.

Marozeau, J., 2011. Models of loudness, in: Florentine, M., Popper, A.N., Fay, R.R. (Eds.), Loudness. Springer, New York, pp. 261–284.

Ménard, M., Gallégo, S., Berger-Vachon, C., Collet, L., Thai-Van, H., 2008. Relationship between loudness growth function and auditory steady-state response in normal-hearing subjects. Hearing Research 235, 105–113.

Moore, B.C.J., 2007. Loudness perception and intensity resolution, in: Cochlear hearing loss. 2nd edition. Wiley, Chichester, pp. 93–116.

Moore, B.C.J., 2012. An introduction to the psychology of hearing. 6th edition. Brill, The Netherlands.

Moore, B.C.J., Glasberg, B.R., 1997. A model of loudness perception applied to cochlear hearing loss. Auditory Neuroscience 3, 289–311.

Müller, J., Janssen, T., 2004. Similarity in loudness and distortion product otoacoustic emission input/output functions: implications for an objective hearing aid adjustment. The Journal of the Acoustical Society of America 115, 3081–3091.

Neely, S.T., Gorga, M.P., Dorn, P.A., 2003. Cochlear compression estimates from measurements of distortion-product otoacoustic emissions. The Journal of the Acoustical Society of America 114, 1499–1507.

Oldfield, R.C., 1971. The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9, 97–113.

Papakonstantinou, A., Kollmeier, B., Riedel, H., 2013. Ipsi- and contralateral interaction in the 40 Hz auditory steady state responses (ASSRs) with two carriers at 60 dB SPL. International Journal of Audiology 52, 626–635.

Picton, T.W., 2011. Auditory steady-state and following responses: dancing to the rhythms, in: Human auditory evoked potentials.. Plural Publishing, San Diego, pp. 285–333.

Picton, T.W., Dimitrijevic, A., Perez-Abalo, M..C., Van Roon, P., 2005. Estimating audiometric thresholds using auditory steady-state responses. Journal of the American Academy of Audiology 16, 140–156.

Picton, T.W., van Roon, P., John, M.S., 2007. Human auditory steady-state responses during sweeps of intensity. Ear and Hearing 28, 542–557.

Pratt, H., Sohmer, H., 1977. Correlations between psychophysical magnitude estimates and simultaneously obtained auditory nerve, brain stem and cortical responses to click stimuli in man. Electroencephalography and Clinical Neurophysiology 43, 802–812.

Rasetshwane, D.M., Neely, S.T., Kopun, J.G., Gorga, M.P., 2013. Relation of distortion-product otoacoustic emission input-output functions to loudness. The Journal of the Acoustical Society of America 134, 369–383.

Reyes, S.A., Lockwood, A.H., Salvi, R.J., Coad, M.L., Wack, D.S., Burkard, R.F., 2005. Mapping the 40-Hz auditory steady-state response using current density reconstructions. Hearing Research 204, 1–15.

Robinson, K, Gatehouse, S, 1996. Test-retest reliability of loudness scaling. Ear and Hearing 17, 120–123.

Röhl, M., Kollmeier, B., Uppenkamp, S., 2011. Spectral loudness summation takes place in the primary auditory cortex. Human Brain Mapping 32, 1483–1496.

Röhl, M., Uppenkamp, S., 2012. Neural coding of sound intensity and loudness in the human auditory system. Journal of the Association for Research in Otolaryngology 13, 369–379.

Serpanos, Y.C., O'Malley, H., Gravel, J.S., 1997. The relationship between loudness intensity functions and the click-ABR wave V latency. Ear and Hearing 18, 409–419.

Silva, I., Epstein, M., 2010. Estimating loudness growth from tone-burst evoked responses. The Journal of the Acoustical Society of America 127, 3629–3642.

Silva, I., Epstein, M., 2012. Objective estimation of loudness growth in hearing-impaired listeners. The Journal of the Acoustical Society of America 131, 353–362.

Steinmann, I., Gutschalk, A., 2011. Potential fMRI correlates of 40-Hz phase locking in primary auditory cortex, thalamus and midbrain. NeuroImage 54, 495–504.

Svensson, E., 2000. Comparison of the quality of assessments using continuous and discrete ordinal rating scales. Biometrical Journal 42, 417–434.

Thorson, M.J., Kopun, J.G., Neely, S.T., Tan, H., Gorga, M.P., 2012. Reliability of distortion-product otoacoustic emissions and their relation to loudness. The Journal of the Acoustical Society of America 131, 1282–1295.

Thwaites, A., Glasberg, B.R., Nimmo-Smith, I., Marslen-Wilson, W.D., Moore, B.C.J., 2016. Representation of instantaneous and short-term loudness in the human cortex. Frontiers in Neuroscience 10, 1–11.

Uppenkamp, S., Röhl, M., 2014. Human auditory neuroimaging of intensity and loudness. Hearing Research 307, 65–73.

Valente, M., Van Vliet, D., 1997. The Independent Hearing Aid Fitting Forum (IHAFF) Protocol. Trends in Amplification 2, 6–35.

Van Eeckhoutte, M., Wouters, J., Francart, T., 2015. Loudness adaptation with modulated stimuli. Acta Acustica united with Acustica 101, 1073–1082.

Wilcox, R.R., Granger, D.A., Clark, F., 2013. Modern robust statistical methods: basics with illustrations using psychobiological data. Universal Journal of Psychology 1, 21–31.

Wilson, K.G., Stelmack, R.M., 1982. Power functions of loudness magnitude estimations and auditory brainstem evoked responses. Perception & Psychophysics 31, 561–565.

Zenker Castro, F., Barajas de Prat, J.J., Larumbe Zabala, E., 2008. Loudness and auditory steady-state responses in normal-hearing subjects. International Journal of Audiology 47, 269–275.

Table 1: Details of the hearing impaired participants: sex (M = male, F = female), age (in years), handedness, ear tested, type of hearing loss (HL = hearing loss, SNHL = sensorineural hearing loss, High/Low freq. SNHL = sensorineural hearing loss with more hearing loss in the high or low frequencies), and pure tone average (PTA) are shown, with the latter calculated as the mean threshold in dB HL across 500, 1000, and 2000 Hz.

Number	Sex	Age (years)	Handedness	Ear tested	Type hearing loss	PTA (dB HL)
1	M	80	Right	Right	High freq. SNHL	38
2	F	80	Right	Right	High freq. SNHL	37
3	M	65	Ambidexter	Left	High freq. SNHL	38
4	M	57	Right	Left	High freq. SNHL	45
5	F	62	Right	Right	Flat SNHL	57
6	F	77	Right	Left	High freq. SNHL	28
7	M	74	Right	Right	High freq. SNHL	38
8	M	27	Right	Left	High freq. SNHL	38
9	M	75	Right	Left	High freq. SNHL	35
10	M	71	Right	Left	High freq. SNHL	33
11	F	68	Right	Left	Mixed HL	53
12	M	68	Right	Left	Mixed HL	43
13	F	37	Right	Left	Low freq. SNHL	28
14	M	77	Right	Left	High freq. SNHL	53
15	F	62	Left	Left	Flat SNHL	57

Table 2: Correlation coefficients between the responses obtained during the training and test phases of the behavioral loudness growth measures. P-values are corrected for multiple comparisons based on Holm's method.

	Measure	Carrier frequency	r	p
Normal-hearing	AME	500 Hz	0.97	$<$0.001
		2000 Hz	0.96	$<$0.001
	GRS	500 Hz	0.95	$<$0.001
		2000 Hz	0.95	$<$0.001
Hearing-impaired	AME	500 Hz	0.96	$<$0.001
		2000 Hz	0.95	$<$0.001
	GRS	500 Hz	0.83	$<$0.001
		2000 Hz	0.88	$<$0.001

Table 3: Results of the linear mixed-effects model. The contrast "Beh-ASSR" compares MSE values that contain ASSR data and that contain only behavioral loudness growth. The contrast "Diff-ASSR" compares MSE values for conditions that contain ASSR data.

Factor	Coefficient	t-value	p-value
Intercept	0.015	12.085	< 0.001
Contrast Beh-ASSR	0.0003	0.487	0.629
Contrast Diff-ASSR	-0.0003	-0.209	0.835
Carrier Frequency	-0.004	-2.459	0.017
Participant Type	-0.008	-4.782	< 0.001
Contrast Beh-ASSR x Carrier Frequency	0.001	0.807	0.423
Contrast Diff-ASSR x Carrier Frequency	-0.001	-0.613	0.542
Contrast Beh-ASSR x Participant Type	-0.001	-1.079	0.286
Contrast Diff-ASSR x Participant Type	-0.0002	-0.114	0.910
Carrier Frequency x Participant Type	0.006	2.594	0.012
Contrast Beh-ASSR x Carrier Frequency x Participant Type	-0.001	-0.545	0.588
Contrast Diff-ASSR x Carrier Frequency x Participant Type	-0.0005	-0.163	0.871

Table 4: MSE values found with electrode configuration P9 and P10 for both groups of participants and both carrier frequencies.

Participant Type	Carrier Frequency	MSE Comparison	Median	Mean	Standard Deviation
NH	500 Hz	AME-GRS	0.010	0.020	0.030
		AME-ASSR	0.017	0.019	0.015
		GRS-ASSR	0.018	0.016	0.011
	2000 Hz	AME-GRS	0.009	0.030	0.042
		AME-ASSR	0.013	0.042	0.067
		GRS-ASSR	0.008	0.009	0.005
HI	500 Hz	AME-GRS	0.007	0.009	0.008
		AME-ASSR	0.005	0.007	0.005
		GRS-ASSR	0.004	0.005	0.005
	2000 Hz	AME-GRS	0.011	0.017	0.020
		AME-ASSR	0.013	0.019	0.022
		GRS-ASSR	0.006	0.007	0.004

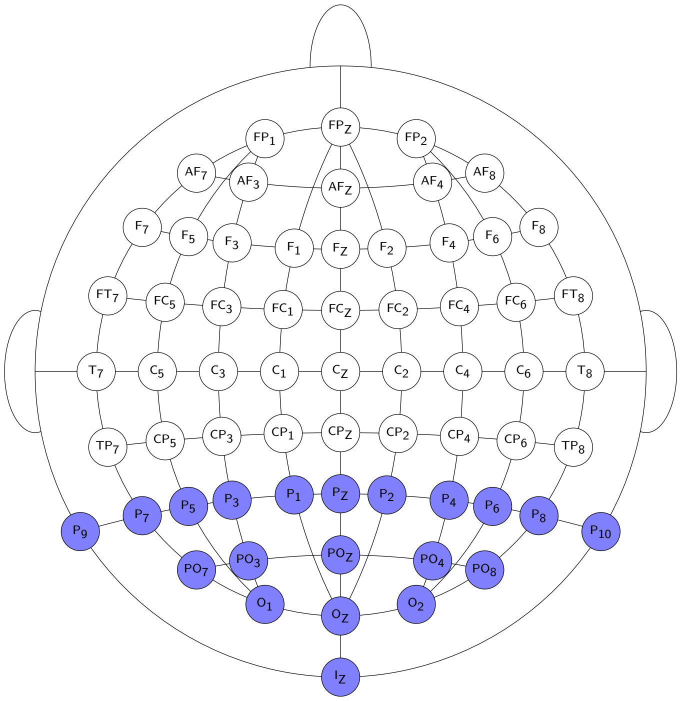

Figure 1: A schematic overview of the 64 Biosemi recording electrodes that were mounted on the head. The electrodes in color were chosen for the ASSR analysis.

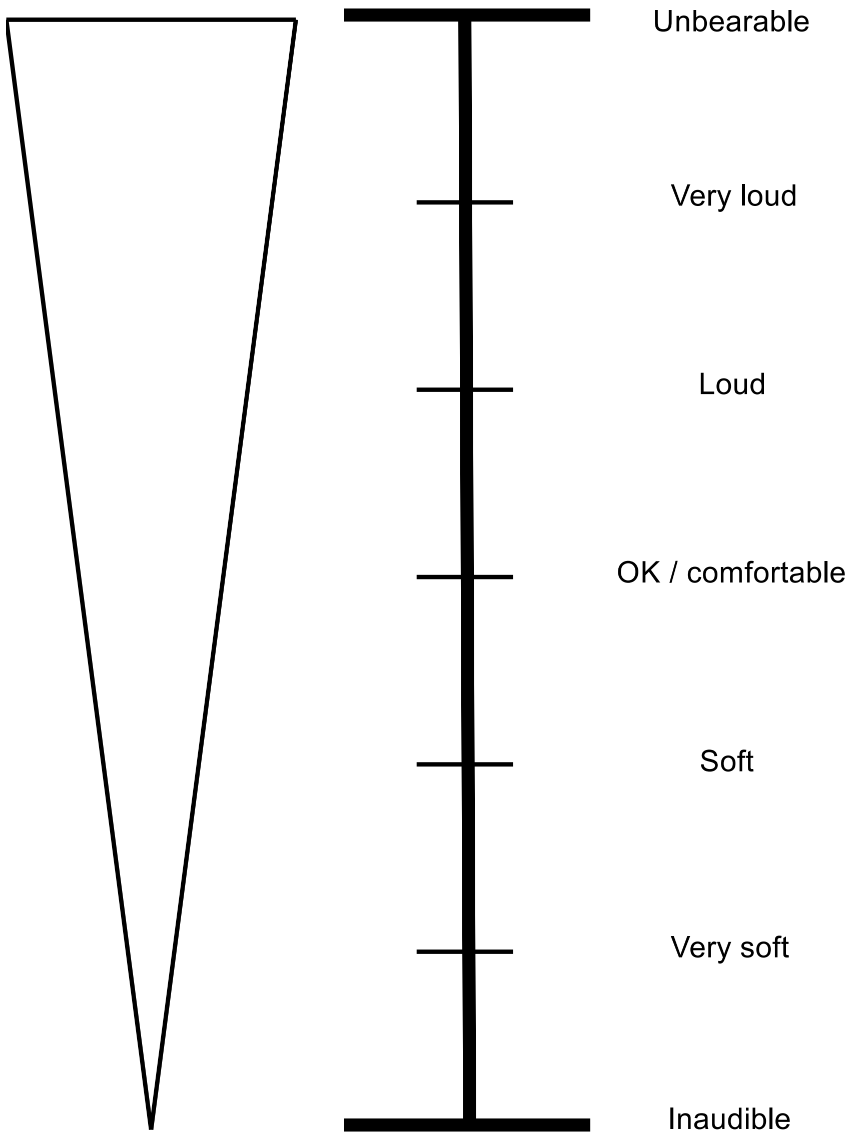

Figure 2: The Graphic Rating Scale used for estimation of behavioral loudness growth functions. The loudness categories were translated from the original labels in Dutch which were: “Onhoorbaar”, “Zeer zacht”, “Zacht”, “OK/comfortabel”, “Luid”, “Zeer luid”, and “Onuitstaanbaar”.

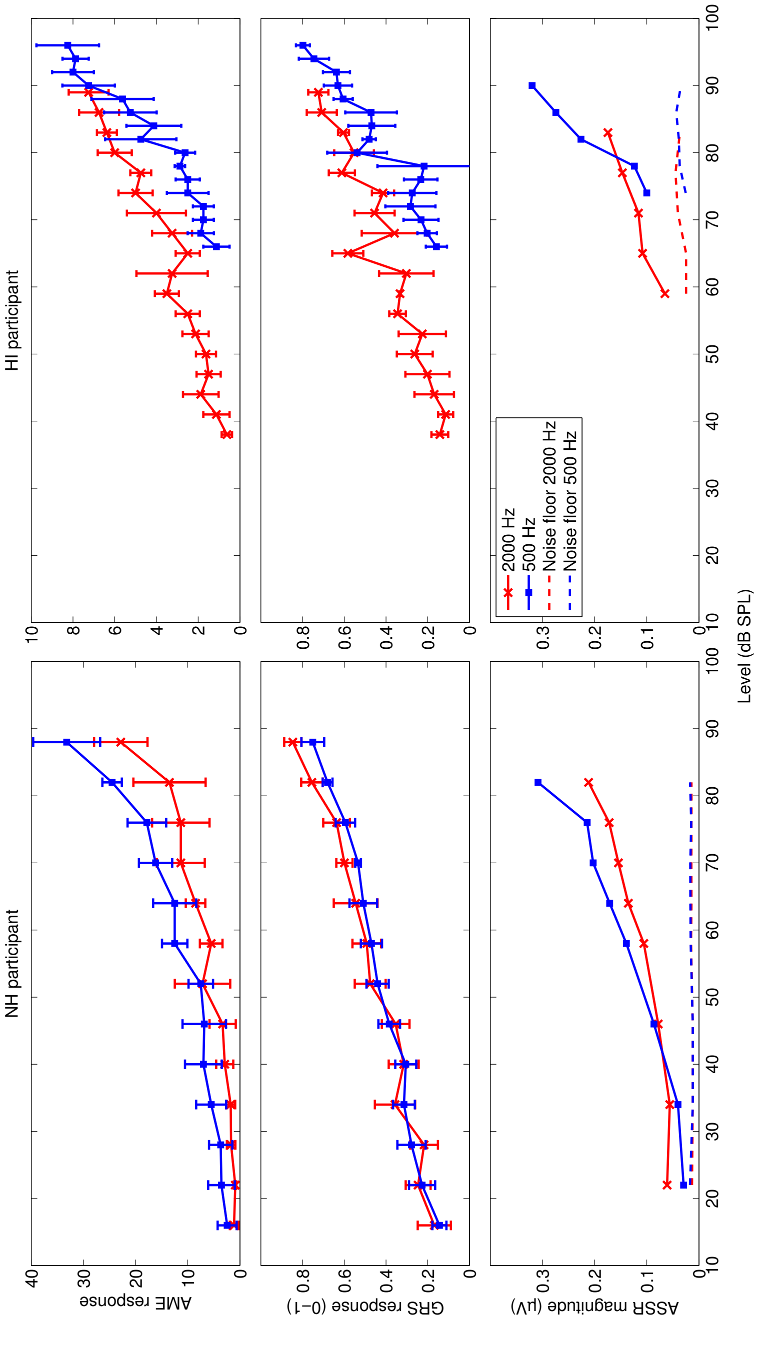

Figure 3: Example results for a normal-hearing and a hearing-impaired participant (number 13). The top panels show the numbers that the participants typed in the Absolute Magnitude Estimation (AME) task (on a linear scale). The middle panels show the responses on the Graphic Rating Scale (GRS). The bottom panels show the ASSR amplitudes (solid lines) and the recorded EEG noise (dashed lines). The 2000 and 500-Hz conditions are indicated with crosses (red in the colored version), and squares (blue in the colored version), respectively. The error bars in the top and middle panels show the mean ± one standard deviation of the responses for each stimulus level.

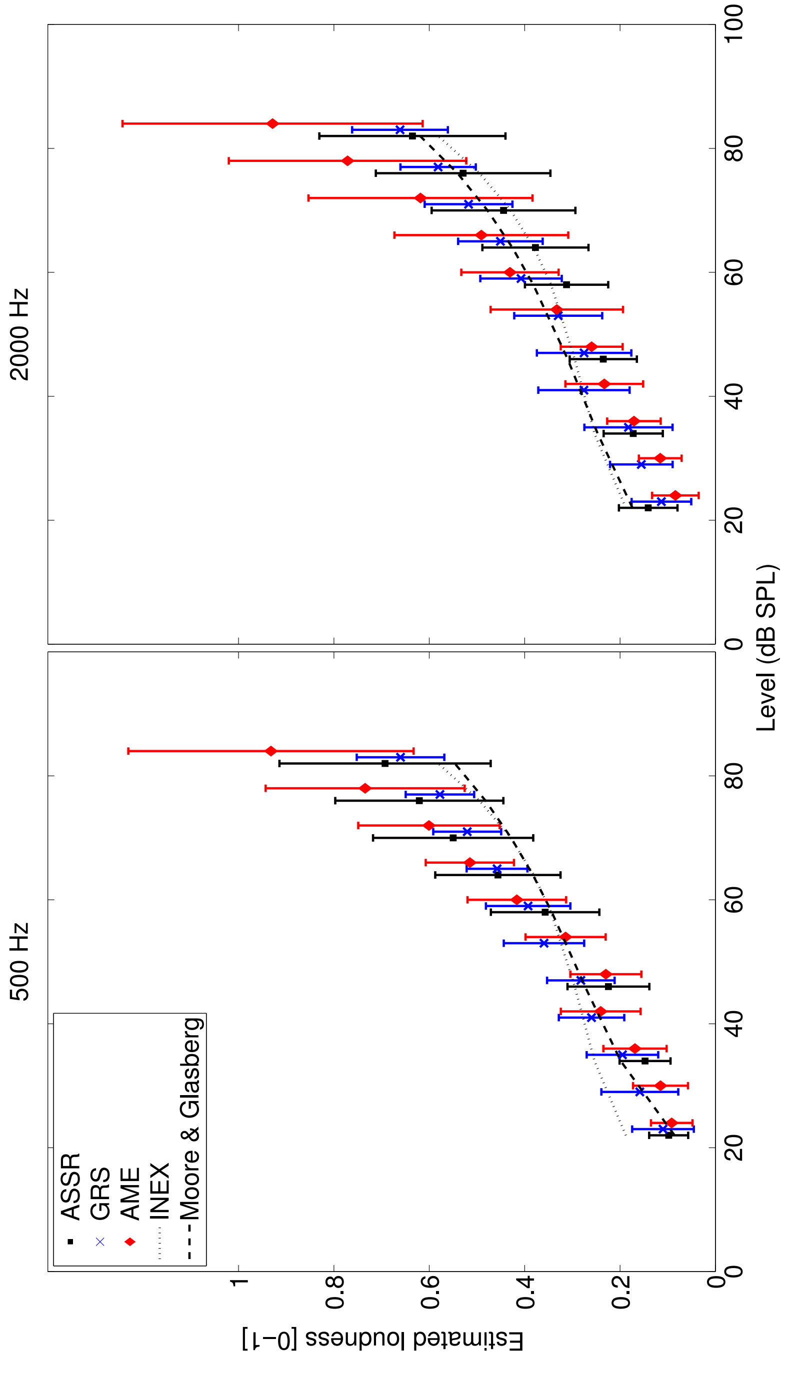

Figure 4: The transformed ASSR and behavioral measures, averaged over all normal-hearing participants, for each carrier frequency. The squares show the ASSR responses, the crosses the GRS responses, and the diamonds the AME responses, and are black, blue and red, respectively, in the colored version. Predictions of two loudness models for normal hearing are plotted on top of the data. Error bars indicate the mean ± one standard deviation.

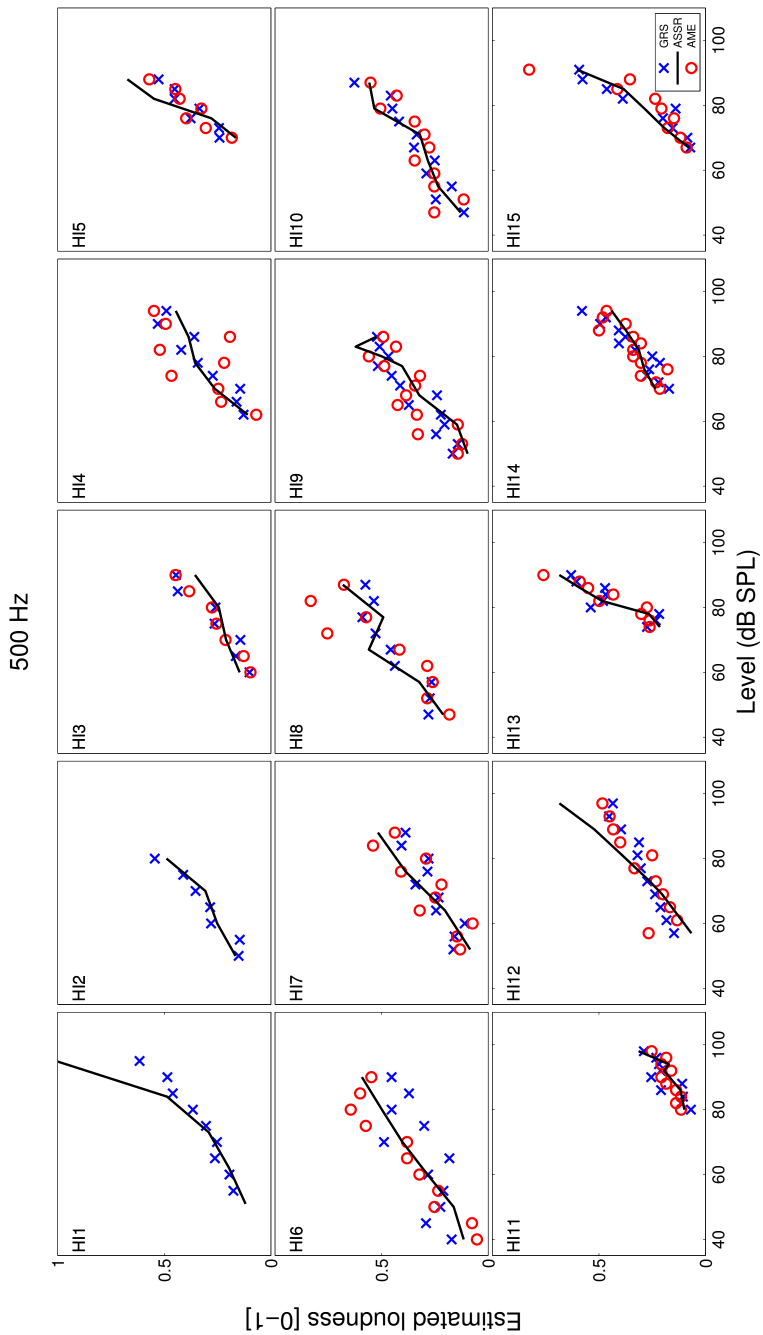

Figure 5: Individual transformed results for the hearing-impaired participants for the 500-Hz stimulus. The behavioral measures GRS and AME are indicated by crosses and circles, and are blue and red in the colored version, respectively. The ASSR responses are indicated by black solid lines.

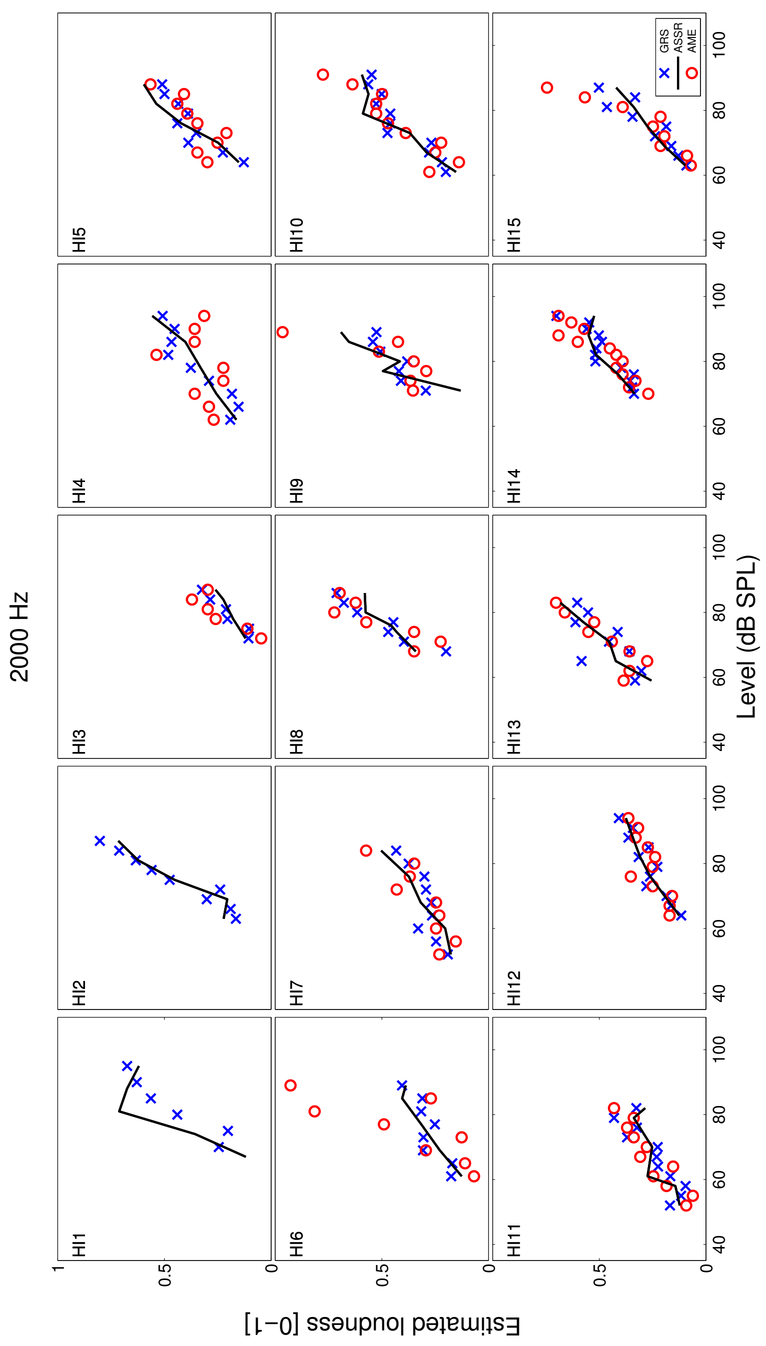

Figure 6: As Figure 5 but for the 2000-Hz stimulus.

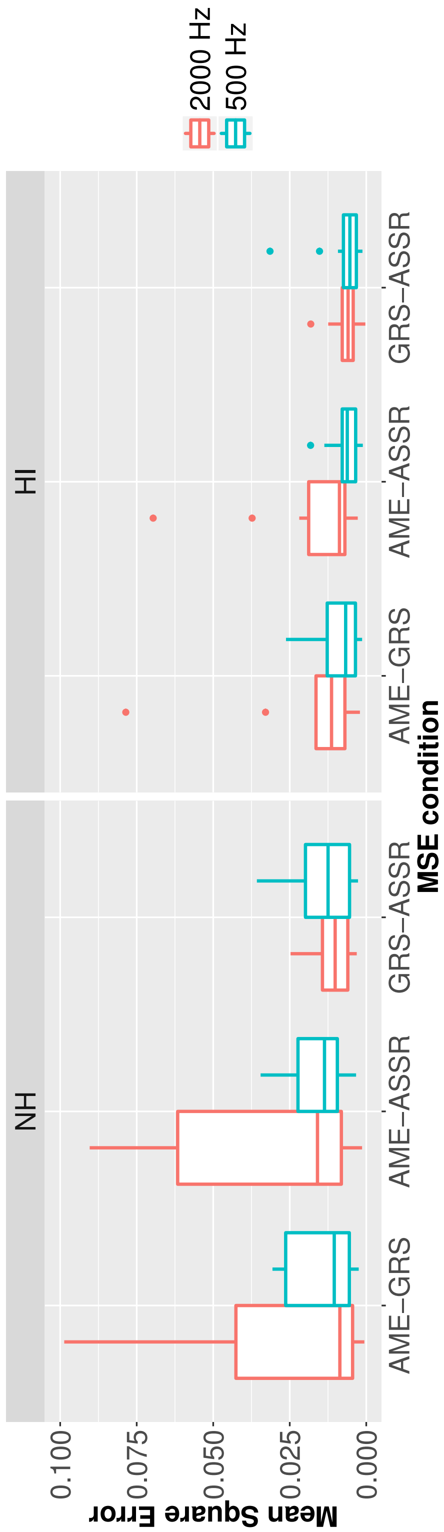

Figure 7: Mean square errors (MSEs) of transformed data for each participant between 1) the AME and GRS responses, 2) the AME and the ASSR responses, and 3) the GRS and the ASSR responses, for the 2000-Hz and 500-Hz carrier frequencies (red and blue in the colored version). For better visibility, 4 outliers of 2 normal-hearing participants were removed: MSE values were 0.14 and 0.28 for the first participant at 2000 Hz and 0.21 and 0.22 for the second participant at 500 Hz, both for MSE comparisons AME-GRS and AME-ASSR.